Technical Note

Black Swans, Fat Tails, and Movie Revenues

Luís Cabral

beta version: September 2009

Coin tosses and the Central Limit Theorem

In 1733, French mathematician Abraham de Moivre did what football referees do best: to toss a coin. In fact, he tossed a coin many times, making note of the outcome after each event. He then grouped the results in series of a certain number of tosses and recorded the percentage of heads in each series. Finally, he plotted the distribution of the percentage of heads after many series of coin tosses.

Boring? Perhaps. But de Moivre was clearly on to something. First he noticed that the distribution of the percentage of heads had a nice regularity about it: a curve shaped like a bell, with many occurrences around 50% and fewer for values closer to 0 or 100%. Second, he got himself thinking about the math underlying this pattern. This led him to postulate a result which later became known as the *Central Limit Theorem:* adding the values of many independent random variables with finite variance results in a variable with an approximately Normal distribution (a.k.a. Gaussian or bell-shaped distribution). The Normal distribution has another beautiful property, known as stability: adding variables that are Normal results in a new variable that is also Normal.

For a long time, few paid attention to de Moivre's idea. Then in 1901 Russian mathematician Aleksandr Lyapunov worked out the math underlying the Central Limit Theorem (the result's name was later given by another mathematician, George Polya). It is hard to underestimate the Theorem's importance in probability theory — many consider it the most important result in the field.[1]

For centuries, the Central Limit Theorem (CLT) has been confirmed by observed distribution patterns; and has provided justification for the use of the Normal distribution in statistical modeling. For example, suppose that the number of tickets sold at each movie theater follows a certain distribution with finite variance; and that the distributions are independent of each other. A movie's box-office revenue is simply the the sum of ticket sales over all of the relevant theaters. If the number of movie theaters is large enough, then movie box-office revenues should be Normally distributed (approximately). It turns out they are not, as we will see next.

Fat tails and un-Normal phenomena

The CLT and all the empirical regularities notwithstanding, there are statistical phenomena with stubbornly un-Normal features. For example, in 1915 economist Wesley Claire Mitchell

---

found that the distribution of stock-price percent changes is not Normal: it includes too few medium-size deviations and too many very small or very large deviations. Two other stochastic processes with very un-Normal distributions include new medical drug revenues and movie revenues.

Exhibit 1 depicts the histogram of US box-office revenues (for all widely distributed movies from 1985–1999). Revenues are cumulative (up to ten weeks after initial release) and expressed in December 1999 US$. Overlaying the histogram is a (truncated) Normal distribution with the same mean and variance as the actual data. As the figure shows, the Normal distribution grossly underestimates the probability of movies with very low revenues and grossly overestimates the probability of movies in the $25–100 million range. There just aren't that many medium- to high-success movies.

Finally, and most important, the Normal distribution grossly underestimates the importance of outliers on the right tail — blockbusters. Since right-tail phenomena are rare, it helps to zoom in the distribution at that level. This is done in Exhibit 2, which shows cumulative distribution functions for earnings greater than $50 million. The cumulative distribution function (cdf) denotes, for a given revenue value, the probability that a movie will make *less* than that value. For example, the Normal distribution predicts that about 96% of all movies make less than $100. The data however shows that only about 94% of all movies make less than $100 million. Or, to put it differently, the probability of a 100+ blockbuster is 6%, not the 4% predicted by the Normal.

The difference between the Normal distribution and the data is most pronounced when we consider values greater than $100 million. For example, the Normal predicts that virtually no blockbuster makes more than $150, but the data shows that between 2 and 3 percent of all movies do so. The biggest blockbuster in the 1985–1999 period, *Titanic*, grossed more than $600 million.

Blockbusters and the Levy distribution

French mathematician Pierre Paul Lévy proposed an alternative to the Normal distribution, one that is now known as the Levy distribution. It is stable, like the Normal, but it does a better job at fitting phenomena with "fat tail" distributions, that is, phenomena with relatively frequent occurrence of very extreme events.

In addition to stock market shifts, Lévy distributions fit well the value of new medical drugs and — you guessed it — movie revenues. Economist Arthur De Vany, who popularized the application to movies, referring to "the long tail to the right, where the blockbusters are located," states that

> These are low probability events of large magnitude that are so far out on the tail they could never occur in a Gaussian world (*Titanic* was about 20 standard deviations above the mean). They can and do occur in the Levy stable world of the movies.[2]

Actually, there are alternative stable worlds (other than the Levy world) where movies could live. One such alternative is the Cauchy distribution. It is hard to pick between Levy and Cauchy when it comes to measuring box-office success. What is quite clear is that the distribution is not Normal.

Black swans and wild uncertainty

In 2007, more than 90 years after Mitchell's original work, NYU professor Nassim Nicholas Taleb published his best-seller, *The Black Swan.* In it, Taleb emphasizes the importance of large magnitude events. He also criticizes the use of the Gaussian distribution as a means to predict future events based on a sample of past events. Although the "black swan theory" includes many elements that go beyond the strictly statistical, the underlying idea is essentially that many phenomena are characterized by distributions with fatter tails than the Normal.

The movie industry is one such phenomenon. De Vany calls it "wild uncertainty." Screenwriter William Goldman rephrases the idea by stating that "nobody knows anything."[3] Former Disney CEO Michael Eisner expands on it: "The movie business has always been like the wild-catting oil business. Everyone wants a gusher."

Endotes

1. Henk Tijms, *Understanding Probability: Chance Rules in Everyday Life,* Cambridge: Cambridge University Press, 2004.

2. http://www.arthurdevany.com/the_movie_business/

3. William Goldman, *Adventures in the Screen Trade*, New York: Warner Books, 1983.

4. Liran Einav, "Seasonality in the U.S. Motion Picture Industry," *Rand Journal of Economics*, 2007.

Exhibit 1

Distribution of US box-office revenues (millions of December 1999 US$).
Source: data cited in Endnote 4 and author's calculations.

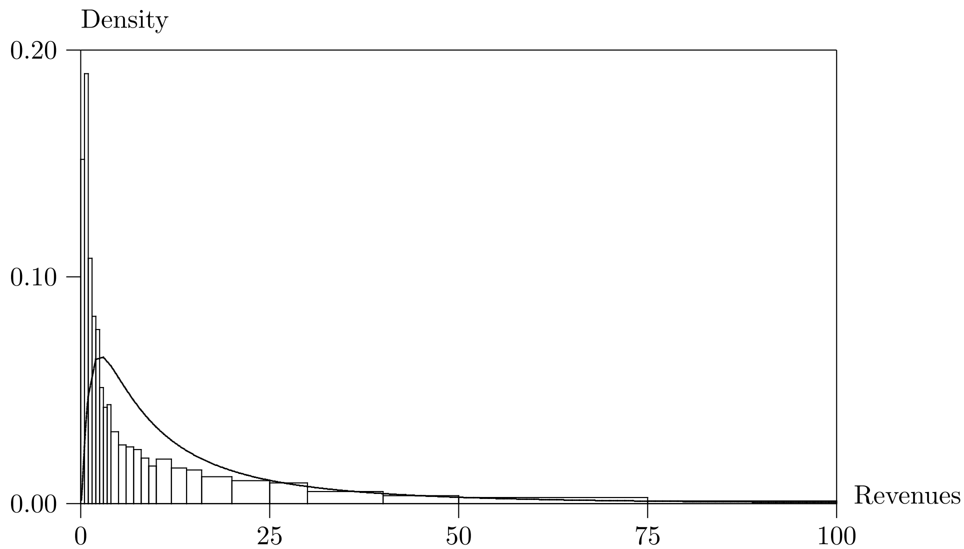

Exhibit 2

Distribution of US box-office revenues (millions of December 1999 US$): blockbusters.
Source: data cited in Endnote 4 and author's calculations.